

## CNES Data Services Activity

# Context

*Part of CNES R&D activity related to EO data systems*

## *Objective :*

- understand the problem of attaching **services** to datasets in addition to **metadata**

## *Way to reach the objective :*

- develop a prototype for testing the resulting concepts on real datasets and services based on:
  - a mechanism for service scheduling and execution (workflow)
  - a catalogue of thematic services

## *Background :*

- experience with the development of a Clearinghouse hosting metadata
- knowledge of work going on within ISO and OGC in the field of services

# User requirement analysis (1/6)

## *Questionnaire*

- circulation of a questionnaire going to a set of representative scientific users of EO data **and** face to face discussions in order to understand their needs and their expectations
- identification of most frequently expected services
- identification of a description model suited to these services
- selection of a first subset of services which could be stored in a “service clearinghouse”

## *Methodology applied*

- get a list of scientists working in the 3 main EO disciplines
  - ocean, atmosphere, solid earth
- get users from various kind of research
  - fundamental, applied, modelling
- get users with various levels of experience in data manipulation
  - manager, senior scientist, junior scientist
- get users from most important research institutes (in France)

# User requirement analysis (2/6)

## *Statistical results*

### ■ level of experience

- ☐ junior 11
- ☐ senior 8
- ☐ manager 4

### ■ discipline

- ☐ atmosphere 4
- ☐ ocean 10
- ☐ solid earth 9

### ■ domain scale

- ☐ global 4
- ☐ local 8
- ☐ meso 3
- ☐ regional 8

### ■ type of research

- ☐ fundamental 2
- ☐ applied 17
- ☐ modelling 4

## User requirement analysis (3/6)

### *Topics covered by the questionnaire*

- domain of activity of the individual
- domain of activity of the group
- EO data used
- data providers
- data products produced
- processes used
- software used
- data difficult to get
- obstacles encountered
- servers and catalogues being used
- example of efficient server or catalogue
- what is missing for data localization in France ?
- is the scientific work hindered by technical constraints on the use of data ?
- expected services
- definition of a “clearinghouse for services”
- efforts scientists are ready to make for the build of such a system
- suggestions

# User requirement analysis (4/6)

## *Expected Services*

## Weighted result

|  |     |
|--|-----|
| ■ availability of software modules for reading data    | 200 |
| ■ data formats descriptions                            | 200 |
| ■ data description                                     | 197 |
| ■ documentation of processing software modules         | 185 |
| ■ availability of processing modules                   | 154 |
| ■ data localization                                    | 144 |
| ■ subsetting   | 138 |
| ■ batch processing for producing higher level products | 132 |
| ■ availability of level 3 products or higher           | 124 |
| ■ supersetting   | 100 |

## *Service Definition*

- scientists had to choose one among four different suggested definitions of a service
  - a small majority (7) chose the most exhaustive and complicated one
  - answers equally spread over the 3 remaining definitions

# User requirement analysis (5/6)

## *Major kinds of processing used*

|                                |    |
|--------------------------------|----|
| ■ atmospheric corrections      | 10 |
| ■ geometric corrections        | 8  |
| ■ radiation transfer inversion | 6  |
| ■ model coupling               | 5  |
| ■ calibration                  | 5  |
| ■ cloud discrimination         | 4  |
| ■ statistic                    | 3  |
| ■ biophysical algorithms       | 2  |
| ■ geophysical corrections      | 2  |
| ■ georeferencing               | 2  |
| ■ temporal series inversion    | 2  |
| ■ quantification algorithms    | 1  |
| ■ data assimilation            | 1  |
| ■ radiometric corrections      | 1  |
| ■ interferometry methods       | 1  |
| ■ multi-sensor                 | 1  |
| ■ pattern recognition          | 1  |
| ■ geodetic transformations     | 1  |

# User requirement analysis (5/6)

## *Software tools used*

|   |   |
|---|---|
| <input type="checkbox"/> ENVI               | 7 |
| <input type="checkbox"/> In house developed | 5 |
| <input type="checkbox"/> IDL                | 5 |
| <input type="checkbox"/> MATLAB             | 5 |
| <input type="checkbox"/> SEADAS             | 4 |
| <input type="checkbox"/> ERDAS Imagine      | 3 |
| <input type="checkbox"/> 6S                 | 2 |
| <input type="checkbox"/> AAPP               | 1 |
| <input type="checkbox"/> ARCV View          | 1 |
| <input type="checkbox"/> BDC                | 1 |
| <input type="checkbox"/> BERNESE (GPS)      | 1 |
| <input type="checkbox"/> ER Mapper          | 1 |
| <input type="checkbox"/> MSPHINX            | 1 |
| <input type="checkbox"/> OS                 | 1 |
| <input type="checkbox"/> PV-Wave            | 1 |
| <input type="checkbox"/> Reading routines   | 1 |
| <input type="checkbox"/> SMAC               | 1 |
| <input type="checkbox"/> TRISKEL            | 1 |



## User requirement analysis (6/6)

*Users say they are ready to :*

- |  |      |
|--|------|
| ■ document their data in their institutes              | 70 % |
| ■ answer questions from a clearinghouse servant        | 70 % |
| ■ reference their data in CNES clearinghouse           | 60 % |
| ■ document their processing modules in their institute | 56 % |
| ■ share their modules with other community members     | 52 % |

## Conclusion

***Most users agree with following definition of a service clearinghouse :***

**“Complete system devoted to localisation, as a metacatalogue allowing subsetting, supersetting, job launching for data level upgrade, and providing previzualization and statistics tools.”**

***Such system should provide following generic services :***

- selection of a geographic area
- selection of a time period
- selection of a thematic
- list of datasets available according to these 3 criteria
- selection of a data subset
- list of available applicable processes
- processing workflow set up
- execution of the processing workflow on a selected datasets
- result exploration (dynamic mapping, statistics)
- retrieval of the processing results

## The way ahead

### *Proceed stepwise*

- implement the CNES metadata clearinghouse
- extend it to support data services
- first implement a limited set of services

### *Look at existing systems*

- DODS ?

### *Stick to existing standards*

- ISO 19119

### *Look at emerging technologies*

- GRID